Received: 21 January 2016 | Revised: 17 April 2017 | Accepted: 5 July 2017

DOI: 10.1111/mec.14241

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

# Extensive seed and pollen dispersal and assortative mating in the rain forest tree *Entandrophragma cylindricum* (Meliaceae) inferred from indirect and direct analyses

Franck Kameni Monthe[1,2] | Olivier J. Hardy[2] | Jean-Louis Doucet[3] | Judy Loo[4] | Jérôme Duminil[1,2,5]

[1]Bioversity International, c/o CIFOR Central Africa Regional Office, Yaoundé, Cameroon

[2]Faculté des Sciences, Service Evolution Biologique et Ecologie, Université Libre de Bruxelles, Bruxelles, Belgium

[3]Foresterie Tropicale, Gestion des Ressources Forestières, BIOSE, Gembloux Agro-Bio Tech, Université de Liège, Gembloux, Belgium

[4]Bioversity International, Rome, Italy

[5]Institut de Recherche pour le Développement, UMR-DIADE, Montpellier, France

**Correspondence**
Franck Kameni Monthe, Faculté des Sciences, Service Evolution Biologique et Ecologie, Université Libre de Bruxelles, Bruxelles, Belgium.
Email: fmonthek@ulb.ac.be

**Funding information**
Congo Basin Forest Fund, Grant/Award Number: P-Z1-C00-024; CGIAR Research Program on Forests, Trees and Agroforestry, Grant/Award Number: P-Z1-C00-024; CGIAR Fund Donors; Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture; Fonds de la Recherche Scientifique

## Abstract

Pollen and seed dispersal are key processes affecting the demographic and evolutionary dynamics of plant species and are also important considerations for the sustainable management of timber trees. Through direct and indirect genetic analyses, we studied the mating system and the extent of pollen and seed dispersal in an economically important timber species, *Entandrophragma cylindricum* (Meliaceae). We genotyped adult trees, seeds and saplings from a 400-ha study plot in a natural forest from East Cameroon using eight nuclear microsatellite markers. The species is mainly outcrossed ($t$ = 0.92), but seeds from the same fruit are often pollinated by the same father (correlated paternity, $r_p$ = 0.77). An average of 4.76 effective pollen donors ($N_{ep}$) per seed tree contributes to the pollination. Seed dispersal was as extensive as pollen dispersal, with a mean dispersal distance in the study plot approaching 600 m, and immigration rates from outside the plot to the central part of the plot reaching 40% for both pollen and seeds. Extensive pollen- and seed-mediated gene flow is further supported by the weak, fine-scale spatial genetic structure ($Sp$ statistic = 0.0058), corresponding to historical gene dispersal distances ($\sigma_g$) reaching approximately 1,500 m. Using an original approach, we showed that the relatedness between mating individuals ($F_{ij}$ = 0.06) was higher than expected by chance, given the extent of pollen dispersal distances (expected $F_{ij}$ = 0.02 according to simulations). This remarkable pattern of assortative mating could be a phenomenon of potentially consequential evolutionary and management significance that deserves to be studied in other plant populations.

KEYWORDS
assortative mating, gene dispersal, mating system, parentage analysis, sustainable management, tropical tree

## 1 | INTRODUCTION

Understanding gene flow patterns of rain forest species is a key requirement for effective species management and conservation, in the light of continuing environmental changes, habitat fragmentation and logging activities (Frankham et al., 2012). Logging generally results in a reduction of population density by eliminating a portion of the largest trees, increasing distance between conspecifics and affecting population regeneration (Cloutier, Kanashiro, Ciampi, & Schoen, 2007; de Lacerda, Kanashiro, & Sebbenn, 2008; Lowe, Boshier, Ward, Bacles, & Navarro, 2005). Hence, seed and pollen dispersal distances, as well as the relationship between fecundity and tree

  

size, are important parameters to be taken into account (Klein, Desassis, & Oddou-Muratorio, 2008; Tani et al., 2012; Thomson, Moles, Auld, & Kingsford, 2011). Trees are often subject to inbreeding depression (Duminil, Hardy, & Petit, 2009; Wang, Hagqvist, & Tigerstedt, 1999), so their reproductive biology and in particular the processes leading to inbreeding (e.g., selfing, assortative mating) must also be well understood to develop sound management strategies of timber trees (Lowe et al., 2005). Although characterizing geneflow patterns has long been challenging (Broquet & Petit, 2009), the development of molecular markers now offers new investigative tools.

Nowadays, to study patterns of gene dispersal within populations, we can distinguish two main categories of molecular-based genetic methods. "Direct" genetic methods provide estimates of contemporary gene flow through the observation of seed and pollen movement, by conducting direct parentage inferences (Marshall, Slate, Kruuk, & Pemberton, 1998) or by modelling forward dispersal kernels (i.e., probability density function of a propagule arrival position around its source position) using parentage probabilities of offspring under a spatially explicit neighbourhood model (Chybicki & Burczyk, 2010). Model-based approaches also provide inference on how reproductive success can be explained by attributes of individuals (e.g., their size) through the analysis of selection gradients (Chybicki & Burczyk, 2010). These methods typically require significant sampling effort and highly polymorphic molecular markers (Ashley, 2010; Gerber, Mariette, Streiff, Bodenes, & Kremer, 2000). Within-population gene dispersal can also be evaluated through "indirect" genetic methods that rely on the spatial genetic structure (SGS) of populations or on patterns of genetic differentiation between the sires of maternal seed families (Robledo-Arnuncio, Austerlitz, & Smouse, 2006; Smouse & Robledo-Arnuncio, 2005; Vekemans & Hardy, 2004). These indirect methods do not rely on direct assignment of offspring to their parents. They allow inferring parameters of backward dispersal kernels (i.e., probability density function of a propagule source position around its arrival position). SGS is often characterized by the relationship between kinship and the spatial distance among individuals within a population, which provides an indication of the mean parent–offspring distance (at migration–drift equilibrium), averaged over several generations (i.e., historical geneflow patterns), without distinguishing the respective role of pollen and seed dispersal (Rousset, 2000; Vekemans & Hardy, 2004). SGS varies widely among forest tree species, notably in relation to species' biological characteristics (Hardy et al., 2006; Vekemans & Hardy, 2004). This approach has been extended to estimate the mean contemporary pollen dispersal distance from the differentiation between the pollen clouds fertilizing different mother trees (Smouse & Sork, 2004).

Assortative mating is the genetic correlation between mating individuals (Fisher, 1918; Wright, 1921). Importantly, flowering phenology has a direct influence on assortative mating as individuals flowering synchronously are more likely to mate (Ison, Wagenius, Reitz, & Ashley, 2014; Weis, Nardone, & Fox, 2014). If phenological variation within population has a genetic basis (Weis & Kossler, 2004), phenological assortative mating implies that genetically related individuals are more likely to mate, leading to biparental inbreeding. Such a pattern can be tested by combining information from (i) a paternity analysis that identifies mates and pollen dispersal distances, and (ii) SGS to predict the expected relatedness between mates conditional on the pollen dispersal distances. If the actual relatedness between mates (i.e., biparental inbreeding of their progeny) is significantly higher or lower than the expected level, a mechanism causing positive or negative assortative mating does occur. Although SGS and limited pollen dispersal have often been invoked to explain signals of biparental inbreeding, to our knowledge, we lack studies quantitatively demonstrating that biparental inbreeding is correctly predicted by the SGS and pollen dispersal distance in plant populations.

Integrating direct and indirect methods can provide new insights. First, comparing historical and contemporary gene dispersal can provide insights on the demographic history of the population. This is because we should expect similar estimates if the population and gene dispersal processes remained stable over several generations (at least in an extensive population of fairly homogeneous density), while recent disturbances should affect mostly contemporary gene dispersal estimates (Hamrick & Trapnell, 2011; Oddou-Muratorio & Klein, 2008). Second, direct and indirect methods can be combined to detect particular biological features such as assortative mating (Weis & Kossler, 2004). In this study, we integrate direct and indirect methods to investigate the patterns of seed and pollen dispersal of an entomophilous and wind-dispersed tropical rain forest tree species, *Entandrophragma cylindricum* Sprague (Meliaceae), in Eastern Cameroon. Our specific objectives are as follows: (i) to characterize the mating system parameters and pollen dispersal distances; (ii) to test whether pollination success is related to the diameter of individual trees; (iii) to characterize the patterns of seed dispersal and to estimate the relative importance of pollen- and seed-mediated gene flow; (iv) to test whether assortative mating occurs; (v) to indirectly evaluate the impact of logging history on gene dispersal patterns in the population via comparison of contemporary and historical gene dispersal. For this purpose, we developed a mixed-sampling strategy that allows analysing the data through direct and indirect approaches.

## 2 | MATERIAL AND METHODS

### 2.1 | Study species

*Entandrophragma cylindricum* (Sprague) Sprague (Meliaceae), known as sapelli, sapele or African mahogany in the timber market, occurs mostly in semi-deciduous forests from eastern Sierra Leone to Uganda (Hall & Swaine, 1976), and is highly exploited in Cameroon (at a minimum diameter of 100 cm) for the quality of its wood. The fruit is a capsule (7–15 cm long, 2–4 cm wide) composed of five individual valves, each containing three to five winged seeds (5–8 cm long, 1–2 cm wide). Hermaphrodite flowers are produced in a loose inflorescence (Palla, Louppe, & Forni, 2002; Styles, 1972). The

species is probably pollinated by insects (bees), and seeds are dispersed by wind (Palla et al., 2002; Styles, 1972). *Entandrophagrama cylindricum* can flower when the diameter at breast height (DBH) reachs 20 cm, but fruits production are generally observed for trees with a DBH > 50 cm (Lourmas, Kjellberg, Dessard, Joly, & Chevallier, 2007; Styles, 1972), although we also found during our field work that the minimum DBH of a fruiting tree was 43.8 cm.

## 2.2 | Sampling

The studied population is localized in Forest Management Units (FMU) 10,046 and 10,060 (3.75–4.33°N, 13.00–14.03°E) in eastern Cameroon (Figure 1). Around these FMU, forest cover is relatively continuous and sapelli is distributed relatively homogenously throughout the region. Selective logging was carried out in the sampled area about 25 years ago, but data on logged species and intensity are not available. The contemporary density of sapelli trees (DBH ≥ 10 cm) was 0.88 ind/ha, and mature trees (DBH ≥ 20 cm) have a mean DBH of 47.5 cm.

To achieve our objectives, we adopted a mixed-sampling strategy by conducting (i) an exhaustive sampling during the fruiting period within a plot to apply direct methods (paternity and parentage analyses) and (ii) nonexhaustive sampling outside the plot at a larger spatial scale to apply indirect methods. More precisely, we first carried out exhaustive sampling of trees and saplings within a 400-ha plot (2 by 2 km), within which we also collected seed families from fruiting trees, particularly in a central 100-ha subplot (Figure 1). In the 400-ha plot, we sampled: 239 matures trees (82 young adults 20 < DBH < 40 cm and 157 adults DBH > 40 cm), 114 saplings (DBH < 20 cm; 56 coming from the 100-ha subplot) and 484 seeds coming from 39 fruiting trees (338 seeds from 27 trees inside the 100-ha subplot). Hence, 49% of the saplings and 70% of the seeds sampled came from the central subplot. In this plot, fruiting was not observed on trees with DBH < 40 cm. Seeds were collected on the ground, below trees, at the beginning of the fruiting season. Some of these seeds were still within fallen fruits, and we kept this information to assess the intrafruit correlated paternity. Second, we conducted nonexhaustive sampling along four forest trails leading out of the plot, to a distance of up to 6 km from the plot edge, in four different directions (Figure 1). We sampled and geo-referenced 154 mature trees (43 young adults and 111 adults), eight saplings and 288 seeds from 24 seed trees. At least 12 seeds per fruiting tree were collected. For each tree and sapling, leaf (and/or cambium for adult trees) were collected and dried in silica gel for future DNA extraction. Seeds were also dried in silica gel, and we used their cotyledons for DNA extraction.

## 2.3 | DNA extraction and genotyping

DNA was extracted using the NucleoSpin 96 Plant II kit (Macherey-Nagel, Düren, Germany). We genotyped 393 mature trees, 122 saplings and 772 seeds at eight nuclear microsatellites with markers developed by Garcia, Noyer, Risterucci, and Chevallier (2004). The microsatellite loci were amplified in two newly developed multiplexes following the protocol of Micheneau, Dauby, Bourland, Doucet, and Hardy (2011). These multiplexes named "Mix I" and "Mix II" were, respectively, composed of six (PEcCIR 12, 22, 156, 227, 244 and 247) and four (PEcCIR 84, 103, 143 and 271) microsatellite markers. PCR amplifications were performed using the Type-it Microsatellite PCR Kit (QIAGEN, Valencia, CA, USA) as follows: 1 μl DNA, 7.5 μl of Multiplex Master Mix, 3.3 μl or 2.2 μl, respectively, for Mix I and II. The PCR volume was adjusted to a final volume of 15 μl with ultrapure water. PCR was run on a TProfessional Basic Thermocycler (Biometra, Göttingen, Germany) using the following cycling condition: initial denaturation at 95°C for 5 min, then 22 cycles at 95°C for 30 s, 57°C for 90 s and 72°C for 30 s, followed by 10 cycles at 95°C for 30 s, 57°C for 90 s and 72°C for 30 s, and a final extension at 60°C for 30 min. For genotyping, a mix of 0.8 μl of PCR product, 12 μl of Hi-Di formamide (Life Technologies) and 0.3 μl of GeneScan *500 LIZ* size standard (Applied Biosystems) were used. Samples were analysed on a 48-capillary 3730×l DNA Analyzer (Applied Biosystems). The sizes of the different fragments were determined using Peak Scanner v1.0 (Applied Biosystems). To evaluate the quality of genotype scoring, we genotyped a random sample of 25 individuals twice.

## 2.4 | Genetic diversity and mating system analyses

For each cohort (adults, saplings, seeds), we computed the mean effective number of alleles ($NA_E$, Nielsen, Tarpy, & Reeve, 2003), the standardized allelic richness, expressed as the expected number of alleles among ($k$ = 212) gene copies ($A_R$), the observed heterozygosity ($H_O$), the unbiased expected heterozygosity ($H_E$, Nei's index of genetic diversity; Nei, 1978) and the fixation index $F$ ($F = 1-H_O/H_E$) using SPAGEDI 1-5a (Hardy & Vekemans, 2002). Because $F$ can be biased by the presence of null alleles, we tested for their presence and estimated Wright's inbreeding coefficient $F_c$ for each cohort, using a method accounting for null alleles and implemented in INEST 1.0 (Chybicki & Burczyk, 2009). Using MLTR 3.2 (Ritland, 2002), we estimated for sampled seed trees: (i) the multilocus outcrossing rate ($t_m$); (ii) the average outcrossing rate for each locus ($t_s$); (iii) the correlated paternity ($r_p$, or the fraction of outcrossed progeny which share both parents); (iv) the correlation of selfing among families ($r_s$), which reflects variation in selfing rates among families and; (v) the rate of biparental inbreeding ($t_m-t_s$), which reflects the degree of mating between relatives. MLTR is based on a maximum-likelihood method. 95% confidence intervals were assessed by bootstrap (1,000 replicates) using families as resampling units.

## 2.5 | Characterization of contemporary geneflow patterns through direct analyses

### 2.5.1 | Identifying pollen dispersal events through parentage analyses

We used likelihood methods to conducted maternity and paternity analyses, using CERVUS 3.0.3 (Marshall et al., 1998). The potential

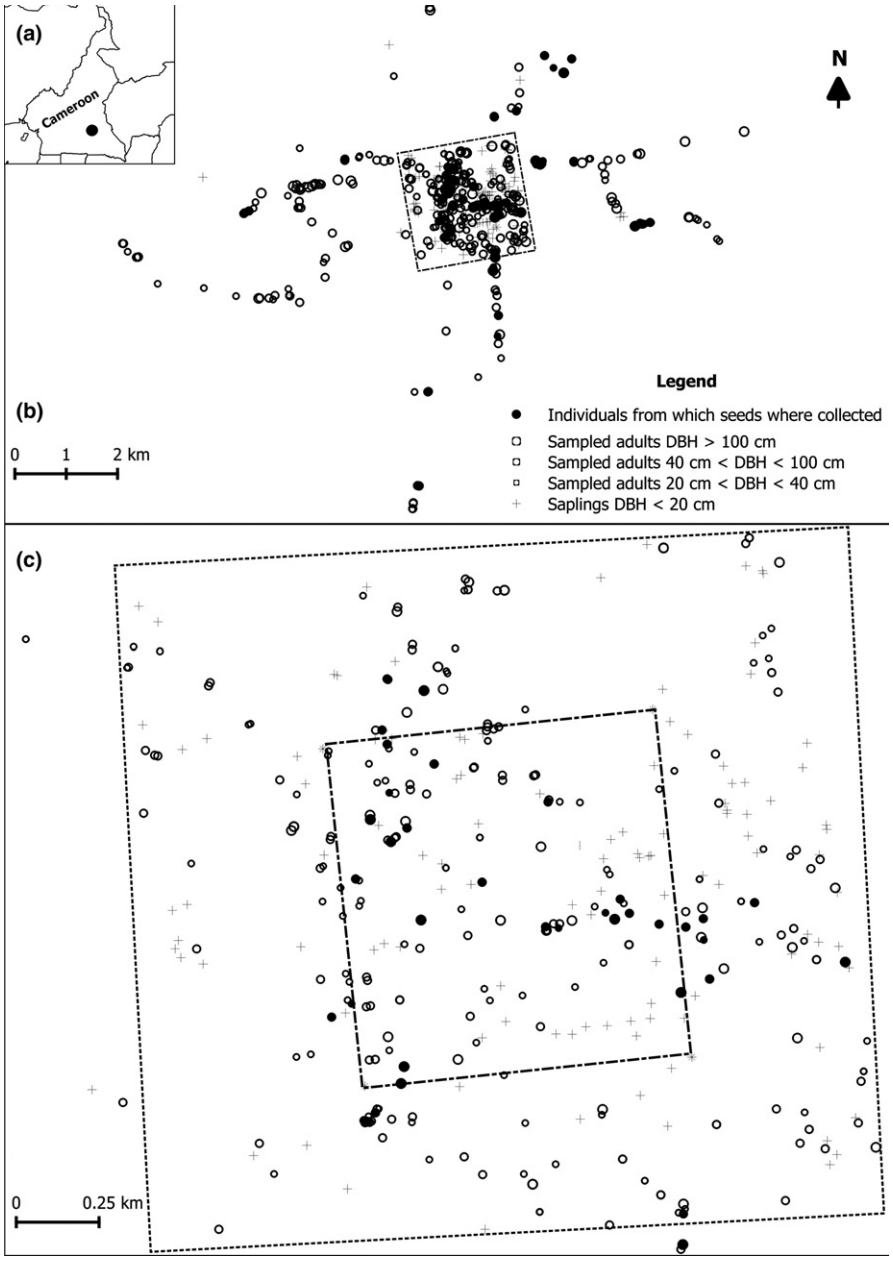


**FIGURE 1** Maps of the study site and sampling strategy. (a) The inset shows the position of the area within Cameroon. (b) Sampling area with the 400-ha exhaustively sampled plot delimited by the dashed line and nonexhaustive sampling along trails surrounding the plot. (c) The external dashed lined square represents the 400-ha plot. The internal square represents a 100-ha central area where efforts to harvest seeds were concentrated. The size of the circles represents DBH sizes

exclusion power of our markers for these analyses was checked using the nonexclusion probability estimated for the first parent. For inferring pollen dispersal events, we first conducted a maternity analysis to test whether seeds collected below mother trees ($n$ = 772) were actually produced by the expected mother. To this end, we relied on a simulation to estimate the critical values of LOD score (log of the odds ratio) required to assign maternity at 80% and 95% confidence levels. The simulation was performed by CERVUS on 10,000 offspring, setting the proportion of sampled candidate mothers as 0.5 (because preliminary analyses indicated that nearly half of seeds were immigrants), the proportion of loci typed at 0.90 (as observed in our data set) and the proportion of mistyped loci at 0.10. We then conducted a paternity analysis at the scale of the 400 ha plot, considering seeds with the expected mother confirmed and all trees from the plot with a DBH ≥ 20 cm as potential pollen donors (mature trees; $n$ = 239). As selfing is possible, mother trees were also considered as candidate fathers. As above, to define the critical LOD scores (at 80% or 95% confidence), we simulated paternity analysis for 10,000 offspring setting the proportion of sampled fathers at 0.50, the proportion of loci typed at 0.90 and the proportion of loci mistyped at 0.10. We then conducted the paternity analysis on all 484 seeds, as they were well assigned by the previous maternity analysis to the expected mother tree. We assigned paternity based on the LOD score of the most likely pollen donor, comparing the multilocus genotype of each candidate father to that of seeds, knowing the mother, as described by Marshall et al. (1998). Likelihood ratios were calculated using the equations of Kalinowski, Taper, and Marshall (2007). The selfing rate was estimated as the proportion of seeds for which maternity and paternity were assigned to the same candidate parent.

### 2.5.2 | Estimating seed and pollen forward dispersal kernels from a spatially explicit neighbourhood model

We applied the maximum-likelihood method implemented in NM+ software (Chybicki & Burczyk, 2010) to estimate: *s*, the proportion of selfing; $m_p$, the proportion of pollen immigration coming from outside the plot; $d_p$, the mean pollen dispersal distance within the plot; $m_s$, the proportion of seeds coming from outside the plot; $d_s$, the mean seed dispersal distance within the plot. To estimate pollen or seed immigrating from outside the 400-ha plot, we relied (i) on adults, saplings and seeds sampled within the 400-ha plot; (ii) on saplings and seeds sampled only within the central 100-ha subplot (allowing us to estimate migration rates corresponding to dispersal events >500 m, which is the minimal distance between the limit of the 100-ha subplot and the limit of the 400-ha plot). Seed and pollen were assumed to disperse following exponential kernels because preliminary analyses showed that data at the scale of the 400-ha plot were not powerful enough to estimate precisely the shape of dispersal kernel functions based on two parameters (e.g., exponential power function). Genotyping error rates (including the effect of null alleles) were estimated by NM+ from the mismatches observed between seeds and maternal genotypes. We first ran NM+ on the seeds to estimate pollen dispersal parameters (s, $m_p$, $d_p$) together with error rates per locus. Fixing these parameters, we then ran NM+ considering the sapling genotypes to estimate seed dispersal parameters ($m_s$, $d_s$). To assess the sensitivity of these parameter estimates to the assumed genotyping error rates, we repeated the procedure while fixing the error rate to 0.01, 0.02, 0.05, 0.10 or 0.15 (same value for all loci). Finally, we tested whether the DBH of candidate parents affected the probability of being a true mother or father using centred and normalized DBH as a selection gradient in NM+ and considering both linear and quadratic effects. We further compared the distribution of DBH of the fathers detected in the previous paternity analysis (CERVUS) with the distribution of DBH of the adult population of the plot.

## 2.6 | Characterization of geneflow patterns through indirect analyses

### 2.6.1 | Estimating historical backward gene dispersal from SGS analyses

Spatial genetic structure was assessed for all adult trees ($n$ = 393) in the population following the procedure described in Vekemans and Hardy (2004). Pairwise kinship coefficients ($F_{ij}$) were estimated between individuals using the Nason's estimator of kinship coefficient (Loiselle, Sork, Nason, & Graham, 1995) as implemented in the software SPAGeDi 1-5a (Hardy & Vekemans, 2002). This estimator has been chosen due to its statistically robust properties (Vekemans & Hardy, 2004). To analyse the SGS, $F_{ij}$ values were regressed on the logarithm of the spatial distances between individuals, $\ln(d_{ij})$, giving the regression slope *b*. Furthermore, to illustrate the SGS graphically, $F_{ij}$ values were averaged over a set of eight nonoverlapping geographical distance intervals (upper distance of each interval in metres: 50, 100, 200, 400, 800, 1,600, 3,200, 6,400). We tested SGS by permuting 10,000 times the spatial position of individuals to obtain the frequency distribution of the regression slope (*b*) under a null hypothesis of no SGS. To quantify the strength of SGS, the *Sp statistic* (Vekemans & Hardy, 2004) was computed as $Sp = -b/(1-F_1)$ where $F_1$ is the mean pairwise kinship coefficient between individuals separated by <50 m, and considered as neighbours.

Finally, assuming drift–dispersal equilibrium, we estimated a characteristic gene dispersal distance, $\sigma_g$ ($\sigma_g^2$ is half the mean-squared parent–offspring distance), and the neighbourhood size, $N_b = 4\pi d_e.\sigma_g^2$ where $d_e$ is an effective population density of reproductive individuals. For this purpose, we used an iterative procedure based on the restricted regression slope ($b_r$) of $F_{ij}$ on $\ln(d_{ij})$ within a limited distance range $\sigma_g > d_{ij} > 20\sigma_g$, and the theoretical expectation that $N_b = -(1-F_1)/b_r$ under isolation by distance, following Vekemans and Hardy (2004). The observed density $d_{obs}$ (density of adult individuals, DBH > 20 cm) in the 400-ha plot was $d_{obs}$ = 0.597 ind/ha. As $d_e$ should be a fraction of $d_{obs}$ dependent on the variance in lifetime reproductive success among adults, we assumed that $d_e$ ranged from one-half to one-tenth of the value of $d_{obs}$ (Hardy et al., 2006). Accordingly, we estimated $N_b$ and $\sigma_g$ for different values of effective densities ($d_{obs}/2$, $d_{obs}/4$ and $d_{obs}/10$).

### 2.6.2 | Estimating correlated paternity and backward pollen dispersal from the spatial structure of pollen pools

Based on genotypes of geo-referenced maternal seed families (within as well as outside the 400-ha plot), we estimated the average within-sibship correlated paternity, the mean pollen dispersal distance and the effective densities, using the KINDIST and TWOGENER methods as implemented in the software POLDISP (Robledo-Arnuncio, Austerlitz, & Smouse, 2007). Input files have been prepared following the recommendations of POLDISP user's manual: (i) no missing data in the mother (if any, we reconstructed the mother genotype from the offspring genotypes); (ii) no mismatch between the genotype of the mother and the offspring (possible mismatches, generally due to null alleles, were coded as missing data in offspring genotypes); (iii) no offspring resulting from selfing (selfed seeds inferred from CERVUS were removed); (iv) a minimum of two seeds per mother. Using KINDIST, we first tested whether seeds from the same fruit were resulting from a unique pollination event (i.e., have the same father). To this end, each fruit was defined as an independent family to estimate the average within-fruit correlated paternity. As the latter was high (see results), we have then kept only one seed per fruit to estimate the correlated paternity within mother trees (between fruits) and between mother trees, ensuring that correlated paternity results from independent dispersal events. We tested whether the among-sibship correlated paternity was inversely correlated with the distance between mother trees using a Mantel test procedure, implemented in the ZT software (Bonnet & Van de Peer, 2002). We estimated the effective number of pollen donors, $N_{ep}$, from the within-sibship correlated paternity ($r_p$) as $N_{ep} = 1/r_p$.

We then tested the fit of the different dispersal kernels available in KINDIST, including the exponential power kernel. As a portion of the available seed families being spread beyond the 400-ha plot, we expected more information from this data set to infer the shape of the pollen dispersal kernel than with the paternity analysis (NM+), justifying the use of two-parameter kernels in the present analysis. In KINDIST, we used 2,000 m as a reference distance threshold to define unrelated pollen pools because the rate of correlated paternity between mother trees did not decay further beyond this distance. Finally, we used TWOGENER to estimate the effective male population density ($D_{Ep}$) using as input file the previous pollen dispersal distribution parameters (dispersal distance and effective density) estimated with KINDIST.

#### 2.6.3 | Integrating direct and indirect methods to test assortative mating

We tested whether mating events between related individuals occurred more, or less often than would be expected by chance, given the spatial extent of pollen dispersal. This analysis was performed considering only the mating events demonstrated earlier by CERVUS at a 95% confidence level ($n$ = 164). Among these 164 mating events, some occurred repeatedly between the same mother–father pair. To avoid replicated observations, we kept only one observation per mother–father pair ($n$ = 91). Using SPAGEDI 1-5a (Vekemans & Hardy, 2004), we calculated the pairwise kinship coefficient $F_{ij}$ between the father and the mother of each of the 91 independent mating events (mean observed $F_{ij}$ between mates). We then computed the expected $F_{ij}$ value for these 91 mating pairs considering only the spatial distance separating mother and father (assuming no form of assortative mating). To this end, for each mating pair, we considered the mean kinship between adults separated by the distance interval including the actual distance between mates, and we averaged the resulting $F_{ij}$ values over the 91 mating pairs (expected $F_{ij}$ between mates). To further test whether the mean observed $F_{ij}$ between mates differs from the expected value conditional on pollen dispersal distance, we designed the following simulation procedure. For each of the 91 mother–father pairs, we selected randomly an adult from the 400-ha plot situated at a distance from the mother of between 0.9 and 1.1 times the mother–father distance. We then computed the mean $F_{ij}$ between the mother and the randomly selected adult for all the 91 pairs, and repeated the random sampling 1,000 times to obtain a distribution of mean $F_{ij}$ that was compared to the mean observed $F_{ij}$ between mates.

Finally, to ensure that mean $F_{ij}$ between mates was not biased by the paternity analysis (a possibility if the latter would preferentially detect fathers with a genotype particularly similar or dissimilar to the maternal genotype), we generated 484 seed genotypes under random mating respecting the sampling scheme per mother tree, the null allele frequencies and the pollen immigration and locus error rates estimated by previous analyses. A paternity analysis was then performed on this artificial data set as performed for the real data set, and the mean kinship between inferred mates was computed, expecting a value close to zero (random mating) in the absence of bias.

## 3 | RESULTS

### 3.1 | Genetic diversity and scoring quality

All microsatellite loci exhibited a high level of polymorphism, with an average of seven alleles per locus in the whole population (Table S1). The mean polymorphic information content (PIC) was 0.77, and the combined nonexclusion probability (first parent, i.e., when no parent is known) was 0.0015. We found similar values of expected heterozygosity $H_E$ in the different cohorts. The uncorrected inbreeding coefficient ($F$) was significantly higher than zero in all cohorts, but the estimates corrected for the presence of null alleles were close to zero, though still significantly positive in seeds (Table 1).

For the 25 twice-genotyped individuals, replicated single-locus genotypes were identical in 90% of cases. In 8% of cases, they differed at one allele (generally one replicate appeared heterozygous while the other one appeared homozygous for one of the alleles present in the heterozygous genotype), and in 2% of cases, they differed at two alleles. These figures correspond to a mean genotyping error rate per allele and replicate of 2.8%. However, it varied significantly among loci, from 0 to 10%.

### 3.2 | Mating system analyses

Outcrossing rates estimated by the different methods are on the same order of magnitude, outlining that the species is mainly outcrossed: 0.925 ($SE$ = 0.031) according to the multilocus estimator of MLTR ($t_m$), 0.848 according to the proportion of inferred nonselfing events by CERVUS (58 selfed seeds over a total of 383 fathers found), 0.859 ($SE$ = 0.017) according to NM+ (1–s). MLTR also showed that (i) correlated selfing among families ($r_s$) reached 0.330 ($SE$ = 0.367); (ii) substantial biparental inbreeding was inferred from the difference ($t_m - t_s$) = 0.163 ($SE$ = 0.021); (iii) correlated paternity ($r_p$) within sibships reached 0.242 ($SE$ = 0.030).

### 3.3 | Characterization of geneflow patterns through direct analyses

#### 3.3.1 | Parentage analyses

Our microsatellite loci have a total exclusion probability of 0.99, which indicates effective capacity for identifying parent–offspring pairs. Maternity analyses indicated that 98% of the seeds collected below a tree were correctly assigned to the expected mother tree. Inside the 400-ha plot, the paternity analysis (CERVUS) assigned a father to 164 seeds at a 95% confidence level (CL) and to 28 additional seeds at a 80% CL, representing 39.6% of seeds. Based on these assignments, the average pollen dispersal distance ($\pm SE$) reached 584 $\pm$ 39 m (95% CL) and 715 $\pm$ 35 m (80% CL).

**TABLE 1** Parameters of genetic diversity for the different cohorts of *Entandrophragma cylindricum* in the studied population for eight microsatellite loci

| Cohort | $N^a$ | $NA_E{}^b$ | $A_R{}^c$ | $H_E{}^d$ | $H_O{}^e$ | $F^f$ | $Fc^g$ (*SE*) |
|---|---|---|---|---|---|---|---|
| Adults | 393 | 8.43 | 16.07 | 0.799 | 0.539 | 0.327 | 0.006 (0.010) |
| Saplings | 122 | 7.65 | 15.12 | 0.775 | 0.538 | 0.306 | 0.0004 (0.001) |
| Seeds | 772 | 7.59 | 15.28 | 0.772 | 0.559 | 0.276 | 0.044 (0.010) |
| All samples | 1,287 | 7.92 | 15.84 | 0.783 | 0.551 | 0.297 | |

[a]Sample size.
[b]Effective number of alleles.
[c]Allelic richness ($k$ = 212).
[d]Gene diversity corrected for sample size.
[e]Observed gene diversity.
[f]Inbreeding coefficient (i.e., apparent heterozygote deficit).
[g]Inbreeding coefficient corrected for null alleles under a population inbreeding model (standard errors).

## 3.4 | Modelling seed and pollen dispersal

Using all the 484 seeds (confirmed mothers) and the 114 saplings available in the 400-ha plot, the neighbourhood model (NM+) analyses estimated seed and pollen immigration rates at $m_s$ = 67 ± 5.8% and $m_p$ = 32 ± 2.5%, and mean seed and pollen dispersal distances according to the exponential kernels at $d_s$ = 446 ± 84 ($n$ = 114) and $d_p$ = 506 ± 35 m. When we considered only the subsets of 338 seeds and 56 saplings from the 100-ha central subplot, immigration rates reached $m_s$ = 39 ± 9.5% and $m_p$ = 40 ± 3.2%, and mean dispersal distances $d_s$ = 421 ± 93 m and $d_p$ = 540 ± 53 m. As will be discussed, the different spatial distributions of sampled seedlings (spread over the 400-ha) versus seeds (concentrated in the 100-ha subplot) must be taken into account to interpret the differences between migration rate estimates at the 400-ha and 100-ha scales.

Using all seeds with confirmed mother, NM+ inferred single-locus genotyping error rates (including null alleles) ranging from 1.4% to 14% (mean of 7.4%). A sensitivity analysis showed that estimates of mean seed and pollen dispersal distances were little affected by the assumed error rates, while the estimated selfing rate increased, and immigration rates decreased, when increasing the assumed genotyping error rates (Figure S1). Nevertheless, parameter estimates remained relatively robust when varying error rates from 5% to 10%. We tested further whether the pollen immigration rate varied among seed trees within the plot and found that it ranged from 5 to 85% and tended to increase when approaching the edge of the plot, but the correlation was only marginally significant ($r$ = −.4; $p$ < .09). The selection gradient analysis of NM+ inferred that male fecundity reached a maximum value for trees with a DBH around 110 cm (linear and quadratic coefficients were significant, Figure 2). Similarly, direct paternity analyses (CERVUS) showed that the distribution of DBH of the fathers inferred at a 95% CL for 164 seeds from the plot differed significantly from the DBH distribution of the adult population ($\chi^2$ = 35.92; $p$ < .0001). The 100- to 120-cm-diameter class contributed most to the pollination and trees from this class showed high fecundity. However, even trees in the 20- to 40-cm-diameter class, despite their lower fecundity, contributed significantly to pollination because they were abundant (Figure 2).

## 3.5 | Characterization of geneflow patterns through indirect analyses

### 3.5.1 | SGS analyses

We found a *Sp statistic* value of 0.0058 which denotes a relatively low SGS. The kinship coefficient $F_{ij}$ decays approximately linearly with the logarithm of spatial distance in two phases: (i) at distance <300 m, it decreases rapidly from a value of approximately 0.09 between neighbouring individuals to near 0 at 300 m; (ii) at distances >300 m, it decays at a much lower rate (Figure 3). The gene dispersal distance ($\sigma_g$) estimated at $d_e = d_{obs}/2$ and $d_{obs}/4$ was, respectively, 1078 and 1500 m, corresponding to a Wright's neighbourhood size ($N_b$) of 493 and 477. For lower density ($d_e = d_{obs}/10$), computations of $\sigma_g$ and $N_b$ did not converge.

### 3.5.2 | Spatial structure of pollen pools

The correlated paternity within sibships varied substantially when estimated among seeds within fruits (intrafruit), $r_{p\ (intrafruit)}$ = 0.77, or among fruits (inter-fruits), $r_{p\ (inter\text{-}fruits)}$ = 0.21. Hence, seeds from the same fruit are most often sired by the same father, which is not the case of seeds sampled from different fruits. Correlated paternity between seed trees decreased significantly with spatial distance ($r$ = −.04, $p$ < .01, Mantel test; Figure S2, supporting information), revealing a significant spatial genetic structure of the pollen pools fertilizing different trees. As the effective number of fathers contributing to the pollination of a mother tree is the reverse of the correlated paternity, we found that $N_{ep}$ = 4.76 fathers. We obtained an estimate of mean backward pollen dispersal distance of 1,435 m using the exponential power model (parameters $a$ = 0.0001 and $b$ = 0.1671). Using the same model, we obtained an estimate of effective density (density of trees participating in the fertilization for the given year) of 0.25 ind/ha, which corresponds to a situation where about 37% of adult trees had participated equally in pollination events during the year of observation.

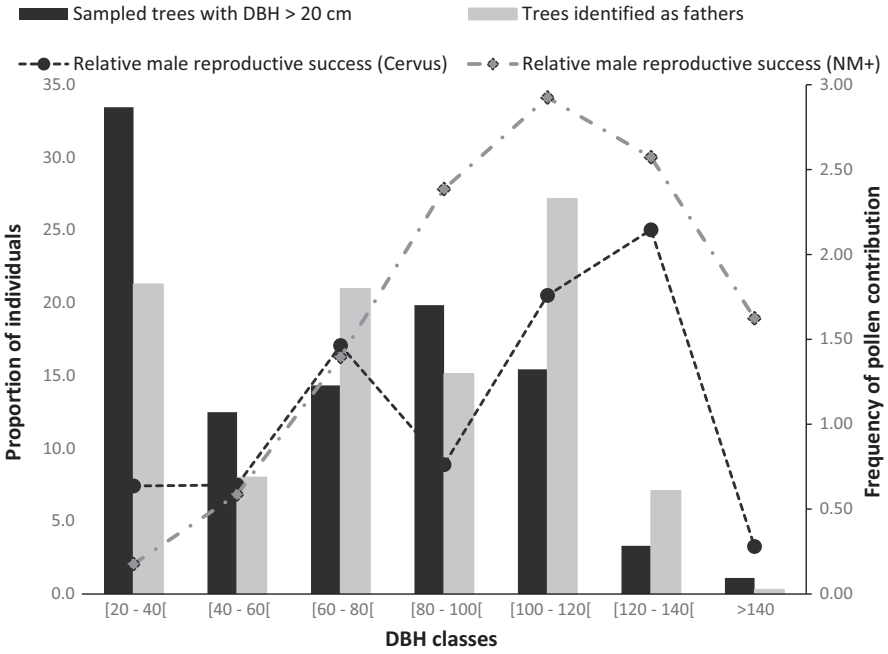


**FIGURE 2** Relationship between tree diameter (DBH) and male reproductive success in the 400-ha study plot. The histograms compare the distributions of DBH of all trees with DBH > 20 cm (black) and the subset of trees detected as father according to a paternity analysis (grey). The curves describe the relative male reproductive success as a function of DBH classes, according to a categorical paternity analysis (CERVUS, circles) or a selection gradient analysis (NM+, diamonds). Maximal male fecundity is found for DBH classes between 100 and 140 cm

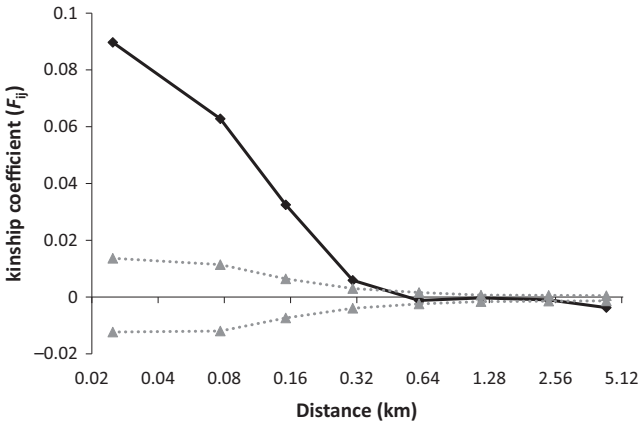


**FIGURE 3** Average kinship coefficients $F_{ij}$ between pairs of adults plotted against the logarithm of geographical distance in the whole population. Dashed lines represent 95% confidence intervals under the null hypothesis that genotypes are randomly distributed

### 3.5.3 | Patterns of mating events: Test of assortative mating

The mean observed kinship between the 91 inferred mating pairs reached $F_{ij}$ = 0.060 (*SE* = 0.013) while the expected mean $F_{ij}$ based solely on their spatial distance and the SGS of adults reached 0.007 (*SE* = 0.002) (Figure 4). When the 91 mating events were simulated under limited pollen dispersal but without assortative mating, the mean $F_{ij}$ between simulated mates was 0.018 (*SD* = 0.010) and over the 1,000 simulation replicates it ranged from −0.009 to 0.051. Hence, an observed value as high as 0.060 was never obtained by simulation ($P < .001$), indicating substantial assortative mating as mating between related individuals occurred more often than expected by chance.

Finally, when we simulated 484 seed genotypes under random mating and performed a paternity analysis on them to identify mates, the mean kinship between these reconstructed mates reached $F_{ij}$ = 0.0017 (*SE* = 0.0006), a value very close to zero. This confirms that our paternity analysis does not identify preferentially genetically related mates and validates our approach to identify assortative mating.

## 4 | DISCUSSION

We have estimated historical and contemporary gene flow in *E. cylindricum* (the African mahogany), an economically important timber species. Hereafter, we first discuss mating system and pollen dispersal characteristics of the species, notably through the comparison of our results with those obtained from studies of the same species by Lourmas et al. (2007), and of other tropical tree species. We discuss the relative importance of pollen- and seed-mediated gene flow in *E. cylindricum*. We then compare our different estimates of gene flow (historical and contemporary) and interpret the evidence for assortative mating. Finally, we discuss our results in the light of logging practices.

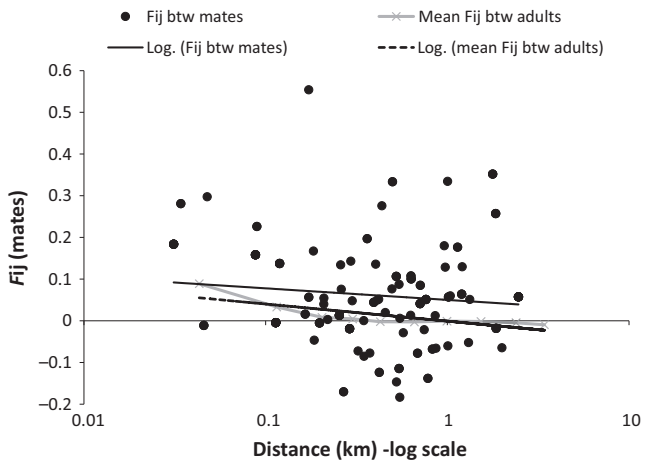


**FIGURE 4** $F_{ij}$ between mates as a function of the distance. The continuous black line corresponds to the trend of the $F_{ij}$ between mates and the dashed black line to the trend of the $F_{ij}$ between adults as obtained from the average $F_{ij}$ between pairs of individuals plotted against distance (grey curve with crosses, which corresponds to Figure 3)

## 4.1 | Comparison of diversity, mating system and geneflow estimates with other studies

The eight microsatellite markers examined in the Cameroonian population of *E. cylindricum* exhibited high levels of genetic diversity with a high number of alleles per locus, allowing us to conduct parentage assignments with high levels of confidence. The power of our markers was however affected by genotyping errors and by the presence of null alleles. This is effectively accounted for by some of the algorithms used, and a sensitivity analysis showed that our dispersal parameter estimates were robust. Parentage analyses were conducted using LOD scores thresholds, assuming fairly high rates of genotyping errors (0.10) compared to the values estimated by genotype replicates (0.028) or mother–offspring mismatches (0.074). In this case, our paternity analyses are likely to be rather conservative and we expect few cases where the inferred father was not the real father.

*E. cylindricum* is predominantly an outcrossed species. Estimates of selfing level diverged slightly between methods (7.5% with MLTR, 14% with NM+ and 15% with CERVUS). The lower estimate obtained with MLTR may be due genotyping errors, not accounted by this algorithm. Lourmas et al. (2007) found lower levels of selfing (<2%). High outcrossing rates have been observed for many tropical forest species (Loveless, 2002; Lowe et al., 2005). We also found that the trace of inbreeding observed within the seed cohort ($F_c$ = 0.044) disappears within the sapling cohort ($F_c$ = 0.0004 ± 0.0013), which may indicate that seeds resulting from selfing die at early stages. This probable counter selection against inbred individuals is a common pattern in long-lived species (Duminil et al., 2009).

*Entandrophragma cylindricum* presents efficient pollen-mediated gene flow. Direct and indirect estimations of pollen dispersal give congruent results. With the neighbourhood model, we estimated that 40% of the pollen contributing to seeds collected within the 100-ha central subplot came from outside the 400-ha plot. Lourmas et al. (2007) obtained 66%–74% pollen immigration, but their seed samples were spread across the entire surface of their different plots, which probably explains a larger contribution of pollen donors from outside the main plot. The relatively long mean distance of within-plot pollen dispersal that we observed (500 m) is also in accordance with Lourmas et al. (2007), who found a mean dispersal distance of 385 m. This capacity for long-distance dispersal can be explained by the behaviour of pollinator insects (bees) as observed in other Meliaceae species (e.g., *Swietenia humilis* and *Cabralea canjerana*; Melo, Coelho, Pereira, Blanco, & Franceschinelli, 2014; White, Boshier, & Powell, 2002). Given the density of mature trees and the long pollen dispersal distance, the degree of connectivity through gene flow might be very high in *E. cylindricum*.

The correlation of paternity ($r_p$), which represents the probability that two seeds have the same father, is closely related to pollination biology (Morgan & Barrett, 1990). The high intrafruit correlation of paternity (0.77) means that most seeds from the same fruit were sired by the same father, probably through a single visit of a pollinator depositing pollen from one source. In contrast, the interfruit correlation of paternity within mothers is much lower (0.21), suggesting that pollen-fertilizing seeds from different fruits of the same tree generally originate from different fathers.

Nevertheless, the effective number of fathers ($N_{ep}$ = 1/0.21 = 4.76) remains low compared to the number of available adult trees within the characteristic distance of pollen dispersal (about 50 adult trees within a radius of 500 m). This may reflect a limitation of the number of available fathers, possibly due to phenological asynchrony, incompatibility between individuals in the population (Sampson, 1998) or unequal pollen fecundity among individuals.

Interestingly, the male reproductive success increases with DBH, a trend also documented by Lourmas et al. (2007) for at least one population, but in our case, it peaks at the 100- to 120-cm-DBH class and declines in larger individuals, as confirmed by the statistical significance of the quadratic term in the selection gradient analysis. Hence, beyond an optimal size, male reproductive success declines, highlighting a senescence process that should be accounted for when forest managers leave a proportion of adult trees for regeneration after exploitation.

Seed dispersal appeared very efficient, in fact as efficient as pollen dispersal. Indeed, although immigration from outside the 400-ha plot was higher for seeds (67%) than for pollen (39%), the difference was essentially due to the fact that most seeds were sampled in the 100-ha central subplot while most saplings were sampled outside it (Figure 1). However, seed and pollen immigration rates became very similar (40%) when only seeds and saplings from the 100-ha central subplot were considered. In *E. cylindricum*, the seed morphology is well adapted to wind dispersal (light, winged seeds) which may allow long-distance dispersal events. Seeds are also possibly secondarily dispersed by animals eating the fruit. By analysing the pollen-to-seed interpopulation migration ratio ($m_p/m_s$) across 93 species through indirect methods, Petit et al. (2005) demonstrated that gene flow by

pollen is predominant (generally by an order of magnitude as they found a median of $m_p/m_s = 17$). However, in 27% of cases, seed dispersal contributed for at least 20% of total gene flow. Among tropical forest trees, the situation in which seed dispersal is as effective as pollen may not be so rare. Some studies of low-density species indicate that seed dispersal could be more effective than pollen dispersal based essentially on the efficiency and morphology of the dispersers, and on floral biology (Bizoux et al., 2009; Born et al., 2008; Duminil, Abessolo et al., 2016; Ndiade-Bourobou et al., 2010).

Historical gene dispersal was also extensive. The estimated *Sp statistic* was 0.0058 denoting a weak spatial genetic structure (Vekemans & Hardy, 2004). This estimation is in line with results observed for other tropical rain forest species pollinated by insects (Dick, Hardy, Jones, & Petit, 2008; Vekemans & Hardy, 2004). Despite a weak SGS, a significant decrease in relatedness between individuals with distance demonstrates that isolation by distance occurs. We observed a rapid decline in relatedness at short distance followed by a slow, approximately linear decline with the logarithm of geographical distance at larger distance, a pattern expected when gene dispersal follows a leptokurtic (also called "fat-tailed") distribution (Hardy et al., 2004; Heuertz, Vekemans, Hausman, Palada, & Hardy, 2003). For instance, this distribution occurs when the seeds are less dispersed than pollen because maternal genes (seeds) disperse at short distance and paternal genes (pollen) at longer distance (Heuertz et al., 2003). In our species, we observed similar dispersal distances for pollen and seeds so that the shape of the kinship curve would be better explained by a leptokurtic dispersal distribution of both seeds and pollen. The estimation of the gene dispersal distance $\sigma_g$ (1,216 m) is also among the highest obtained for tropical wind-dispersed tree species (Hardy et al., 2006). The corresponding large Wright's neighbourhood size ($N_b$ of approximately 500) also suggests a high level of historical gene flow.

## 4.2 | Comparison of historical and contemporary estimates of gene flow

The relatively recent availability of molecular tools has opened a wide range of methodological possibilities (direct, indirect) for the characterization of geneflow patterns (Broquet & Petit, 2009). As the two categories of methods rely on different sampling schemes, they are rarely used together and the vast majority of geneflow studies are based either on direct or indirect methods (Debout, Doucet, & Hardy, 2010; Degen, Bandou, & Caron, 2004; Lowe et al., 2005; Slatkin & Barton, 1989). Here, we were not able to characterize the respective importance of historical pollen- and seed-mediated gene flow, but only obtained an estimate of backward gene dispersal distance of $\sigma_g$ between 1,100 and 1,500 m (according to assumed effective population density), which would correspond to $\sigma_p = \sigma_s$ ranging between 980 and 1,200 m if we assume equal seed ($\sigma_s$) and pollen ($\sigma_p$) dispersal, because $\sigma_g^2 = 1/2\ \sigma_p^2 + \sigma_s^2$. The immigration calculator implemented in NM+ allows relating the proportion of dispersal events going beyond a distance threshold, the neighbourhood radius, with dispersal kernel parameters. Assuming that both seed and pollen dispersal follow an exponential kernel, for a neighbourhood of 1 km radius (which approximates our sampling scheme), a rate of 40% of pollen and seed immigration (as estimated by direct methods) is expected under a dispersal kernel with a mean dispersal distance of 989 m, which translates into $\sigma_p = \sigma_s = 856$ m (Austerlitz et al., 2004), a value slightly lower than our indirect estimates ($\sigma_p = \sigma_s = 980$–1,200 m). Direct and indirect estimates converge even more when assuming a more fat-tailed dispersal kernel, as supported by the KINDIST analysis (where the shape parameter of the power exponential kernel, *b*, is much lower than unity; results not shown). Hence, even though direct methods provided more accurate results than indirect methods, we globally obtained comparable estimates of gene flow. This suggests that the recent history (25 years ago) of logging activities within the study area has not significantly impacted patterns of gene flow. This is not surprising as we have demonstrated that the species is mainly outcrossing and has efficient pollen- and seed-mediated gene flow. We could have expected a larger pollen dispersal distance after the reduction of the population density following logging activities, if nearest-neighbour mating was the rule (see e.g., Duminil, Daïnou et al., 2016). This does not seem to apply here, probably because the species shows large pollen-mediated gene dispersal capacities.

## 4.3 | Biparental inbreeding and assortative mating

Substantial biparental inbreeding was detected by the difference between single-locus and multilocus outcrossing rates (MLTR), contrary to the observations of Lourmas et al. (2007). Biparental inbreeding is expected when pollen dispersal is spatially limited and there is a substantial spatial genetic structure, and the occurrence of such inbreeding in plant populations is often interpreted in these terms but without a quantitative demonstration. However, when computing the expected inbreeding of outcrossed seeds due to limited pollen dispersal in our study population, we found a low value because most pollen disperses over 360 m while the SGS is very limited ($F_{ij} < 0.01$) beyond 300 m. In fact, the mean kinship between the 91 confirmed mating pairs ($F_{ij} = 0.06$) was much larger than the value expected from limited pollen dispersal and SGS ($F_{ij} < 0.01$), implying that a mechanism favouring assortative mating occurs. The mechanism itself is unknown but it may be explained if flowering phenology is governed by strong genetic determinism, so that more related individuals are more likely to be synchronous. Such genetic determinism has been shown in the African rain forest tree *Milicia excelsa* (Daïnou, Doucet, Sinsin, & Mahy, 2012). Alternatively, pollinators' travelling behaviour can possibly be influenced by genetically determined floral traits, such as variation in flower size, shape, colour or chemical cues, which in turn support crossing between individuals sharing similar character states. If our hypotheses proved true, the absence of significant biparental inbreeding in the populations studied by Lourmas et al. (2007) would indicate that some flowering characteristics and/or their heritability may differ among *E. cylindricum* populations. More investigation regarding the determinism of

phenology in tropical trees is needed, given the potentially important evolutionary consequences (Soularue & Kremer, 2014).

More generally, we argue that comparing the kinship between mates with the value expected from pollen dispersal distances and SGS merits being applied systematically when conducting paternity analyses to assess whether assortative mating is more common than previously thought. Presently, it is unclear whether the pattern of assortative mating we detected in *E. cylindricum* is exceptional and specific to the population we studied, or whether it could be common in plant populations but remained unnoticed by lack of adequate method.

### 4.4 | Logging implications

Logging is an important source of income for Congo Basin countries (de Wasseige et al., 2012). However, this activity has long been considered to have negative consequences for the survival of forest tree species, by increasing distances between conspecific trees and modifying abiotic and biotic features of the forest habitat (Lee, 2000; Wickneswari, Ho, Lee, & Lee, 2004). Some studies show that logging activities result in a loss of more than 10% of allelic richness in the logged species although they do not find a negative impact on other genetic diversity parameters (Obayashi et al., 2002; Vinson, Kanashiro, Harris, & Boshier, 2015), possibly because, under genetic drift, the allelic richness of highly polymorphic markers decays much faster than their heterozygosity. Nevertheless, it remains unclear whether short-term observations can be sufficient to detect a significant impact of logging (Lowe et al., 2005).

By comparing historical and contemporary gene flow in a previously logged area, we found no dramatic impact on gene flow. As noted by Lourmas et al. (2007), and confirmed here, 13% of pollen contribution comes from young reproductively mature trees. Moreover, seed production has been observed on trees with a minimum DBH of 43.8 cm. These reproductive features may contribute to ensuring gene flow between individuals while maintaining a number of fathers participating in pollination; thus maintaining genetic diversity in the population. The impact of logging on gene flow in the context of tropical forests is difficult to assess because it depends on the biological and ecological processes specific to each species as well as on the way forests are managed. In Cameroon, management plans usually apply selective logging carried out in blocks that are harvested on a rotation of 30 years, so that the surrounding forest may not yet have been exploited or selectively harvested once in the past 30 years (rotation in FMU exploitation). These unlogged areas and the selective logging may play a key role in the survival of populations by ensuring a constant supply of genes. Also, the extent of pollen-mediated gene flow might increase when population density is reduced (Duminil, Daïnou et al., 2016). Thus in the case of sapelli, long-distance seed and pollen dispersal allow maintenance of diversity. Moreover, being a nonpioneer, light-demanding species, the sapelli might benefit from forest exploitation which creates canopy gaps (Hawthorne, 1995). Hence, disturbances should not be considered as necessarily deleterious to gene flow and genetic diversity for all forest tree species.

## 5 | CONCLUSION

Direct and indirect methods have demonstrated similar patterns of gene flow in *E. cylindricum*. The species exhibits efficient pollen- and seed-mediated gene flow. Such extensive gene flow, combined with large effective population sizes, might contribute to the relatively high genetic diversity observed in the species. As the species is capable of long-distance seed dispersal, it probably has good colonization capacity. Altogether, the demonstration of large pollen dispersal capacities tends to demonstrate that current logging practices will not strongly affect possibilities of cross-pollination in this species. Nevertheless, we have demonstrated that relatively few pollen donors participate in the pollination of seed trees compared to the number of potential mates surrounding them. Moreover, substantial biparental inbreeding was detected and implied relatedness-driven assortative mating. These patterns could be explained by flowering asynchrony among trees in a given year and/or irregular flowering from 1 year to the next, combined with substantial genetic control of phenology. Further investigation of the flowering synchrony among trees is needed as it may have important evolutionary consequences. The original methodology proposed in this study, combining paternity analysis with an analysis of kinship coefficients in space and between mates, can be easily reproduced in other paternity studies to assess the importance of assortative mating in plant populations.

### ACKNOWLEDGEMENTS

This research was supported by the Congo Basin Forest Fund (Beyond Timber project; P-Z1-C00-024), by the CGIAR Research Program on Forests, Trees and Agroforestry and the CGIAR Fund Donors, and by the project "Support to the application of FSC standards on the regeneration and genetic diversity of Congo Basin species" under the Partnership Agreement between COMIFAC and KfW for the Promotion of certified forest exploitation (PPECF). FKM and OJH are funded by the "Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture" (FRIA, Belgium) and the "Fonds de la Recherche Scientifique" (F.R.S-FNRS, Belgium). We thank the SCTB (Société Camerounaise de transformation de bois) and in particular Mr. Roger Embolo for welcoming us during our field work, as well as all the technicians for their help when collecting samples. We are grateful to Peter Smouse and other anonymous reviewers for their constructive comments that helped improve our manuscript.

### DATA ACCESSIBILITY

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.09330.

## AUTHOR CONTRIBUTIONS

O.J.H. and J.D. designed the research; F.K.M. collected samples and generated the data; F.K.M., O.J.H. and J.D. performed the analyses; F.K.M., O.J.H. and J.D. wrote the manuscript; J.L.D. and J.L. commented on the research and on the manuscript, and all authors critically revised the manuscript.

## REFERENCES


Ashley, M. V. (2010). Plant parentage, pollination, and dispersal: How DNA microsatellites have altered the landscape. *Critical Reviews in Plant Sciences*, *29*, 148–161.

Austerlitz, F., Dick, C. W., Dutech, C., Hardy, O., Heuertz, M., Mahy, G., & Doucet, J. L. (2004). Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, *13*, 937–954.

Bizoux, J. P., Daïnou, K., Bourland, N., Hardy, O., Heuertz, M., Mahy, G., & Doucet, J. L. (2009). Spatial genetic structure in *Milicia excelsa* (Moraceae) indicates extensive gene dispersal in a low-density wind-pollinated tropical tree. *Molecular Ecology*, *18*, 4398–4408.

Bonnet, E., & Van de Peer, Y. (2002). zt: A software tool for simple and partial Mantel tests. *Journal of Statistical software*, *7*, 1–12.

Born, C., Hardy, O. J., Chevallier, M. H., Ossari, S., Atteke, C., Wickings, E. J., & Hossaert-Mckey, M. (2008). Small-scale spatial genetic structure in the Central African rainforest tree species *Aucoumea klaineana*: A stepwise approach to infer the impact of limited gene dispersal, population history and habitat fragmentation. *Molecular Ecology*, *17*, 2041–2050.

Broquet, T., & Petit, E. J. (2009). Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 193–216.

Chybicki, I. J., & Burczyk, J. (2009). Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity*, *100*, 106–113.

Chybicki, I. J., & Burczyk, J. (2010). NM+: Software implementing parentage-based models for estimating gene dispersal and mating patterns in plants. *Molecular Ecology*, Resour, *10*, 1071–1075.

Cloutier, D., Kanashiro, M., Ciampi, A. Y., & Schoen, D. J. (2007). Impact of selective logging on inbreeding and gene dispersal in an Amazonian tree population of *Carapa guianensis* Aubl. *Molecular Ecology*, *16*, 797–809.

Daïnou, K., Doucet, J.-L., Sinsin, B., & Mahy, G. (2012). Identité et écologie des espèces forestières commerciales d'Afrique Centrale: Le cas de *Milicia spp*. *Basse*, *16*, 229–241.

de Lacerda, A. E. B., Kanashiro, M., & Sebbenn, A. M. (2008). Effects of reduced impact logging on genetic diversity and spatial genetic structure of a *Hymenaea courbaril* population in the Brazilian Amazon Forest. *Forest Ecology and Management*, *255*, 1034–1043.

de Wasseige, C., De Marcken, P., Bayol, N., Hiol Hiol, F., Mayaux, P., Desclée, B., … Eba'a Atyi, R. (Eds.) (2012). *Les forêts du bassin du Congo Etat des Forêts 2010*. Luxembourg City, Luxembourg: Office des publications de l'Union Européenne. ISBN 978-92-79-22717-2.

Debout, G. D., Doucet, J. L., & Hardy, O. J. (2010). Population history and gene dispersal inferred from spatial genetic structure of a Central African timber tree, *Distemonanthus benthamianus* (Caesalpinioideae). *Heredity* (Edinburgh), *106*, 88–99.

Degen, B., Bandou, E., & Caron, H. (2004). Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana. *Heredity* (Edinburgh), *93*, 585–591.

Dick, C. W., Hardy, O. J., Jones, F. A., & Petit, R. J. (2008). Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. *Tropical Plant Biology*, *1*, 20–33.

Duminil, J., Abessolo, D. M., Bourobou, D. N., Doucet, J.-L., Loo, J., & Hardy, O. J. (2016). High selfing rate, limited pollen dispersal and inbreeding depression in the emblematic African rain forest tree *Baillonella toxisperma*–Management implications. *Forest Ecology and Management*, *379*, 20–29.

Duminil, J., Daïnou, K., Kaviriri, D., Gillet, P., Loo, J., Doucet, J.-L., & Hardy, O. (2016). Relationships between population density, fine-scale genetic structure, mating system and pollen dispersal in a timber tree from African rainforests. *Heredity* (Edinburgh), *116*, 295–303.

Duminil, J., Hardy, O. J., & Petit, R. J. (2009). Plant traits correlated with generation time directly affect inbreeding depression and mating system and indirectly genetic structure. *BMC Evolutionary Biology*, *9*, 177.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, 333–433.

Frankham, R., Ballou, J. D., Dudash, M. R., Eldridge, M. D. B., Fenster, C. B., Lacy, R. C., … Ryder, O. A. (2012). Implications of different species concepts for conserving biodiversity. *Biological Conservation*, *153*, 25–31.

Garcia, F., Noyer, J.-L., Risterucci, A.-M., & Chevallier, M.-H. (2004). Genotyping of mature trees of *Entandrophragma cylindricum* with microsatellites. *Journal of Heredity*, *95*, 454–457.

Gerber, S., Mariette, S., Streiff, R., Bodenes, C., & Kremer, A. (2000). Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Molecular Ecology*, *9*, 1037–1048.

Hall, J., & Swaine, M. D. (1976). Classification and ecology of closed-canopy forest in Ghana. *The Journal of Ecology*, *64*, 913–951.

Hamrick, J., & Trapnell, D. W. (2011). Using population genetic analyses to understand seed dispersal patterns. *Acta Oecologica*, *37*, 641–649.

Hardy, O. J., González-Martínez, S. C., Colas, B., Fréville, H., Mignot, A., & Olivieri, I. (2004). Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae). II. Correlated paternity within and among sibships. *Genetics*, *168*, 1601–1614.

Hardy, O. J., Maggia, L., Bandou, E., Breyne, P., Caron, H., Chevallier, M. H., … Degen, B. (2006). Fine-scale genetic structure and gene dispersal inferences in 10 neotropical tree species. *Molecular Ecology*, *15*, 559–571.

Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, *2*, 618–620.

Hawthorne, W. D. (1995). *Ecological profiles of Ghanaian forest trees*. Tropical Forestry Papers 29. Oxford, UK: Oxford Forestry Institute.

Heuertz, M., Vekemans, X., Hausman, J. F., Palada, M., & Hardy, O. J. (2003). Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology*, *12*, 2483–2495.

Ison, J. L., Wagenius, S., Reitz, D., & Ashley, M. V. (2014). Mating between *Echinacea angustifolia* (Asteraceae) individuals increases with their flowering synchrony and spatial proximity. *American Journal of Botany*, *101*, 180–189.

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, *16*, 1099–1106.

Klein, E., Desassis, N., & Oddou-Muratorio, S. (2008). Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. IV. Whole interindividual variance of male fecundity estimated jointly with the dispersal kernel. *Molecular Ecology*, *17*, 3323–3336.

Lee, S. L. (2000). Mating system parameters of *Dryobalanops aromatica* Gaertn. f. (Dipterocarpaceae) in three different forest types and a seed orchard. *Heredity* (Edinburgh), *85*, 338–345.

Loiselle, B. A., Sork, V. L., Nason, J., & Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, *American Journal of Botany*, *82*, 1420–1425.

Lourmas, M., Kjellberg, F., Dessard, H., Joly, H. I., & Chevallier, M. H. (2007). Reduced density due to logging and its consequences on

mating system and pollen flow in the African mahogany *Entandrophragma cylindricum*. *Heredity* (Edinburgh), *99*, 151–160.

Loveless, M. D. (2002). Genetic diversity and differentiation in tropical trees. In B. Degen, M. D. Loveless & A. Kremer (Eds.), *Modelling and experimental research on genetic processes in tropical and temperate forests* (2000: Kourou, French Guiana) (pp. 3–30), Belem, Brazil: Amazonia Oriental.

Lowe, A., Boshier, D., Ward, M., Bacles, C., & Navarro, C. (2005). Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity* (Edinburgh), *95*, 255–273.

Marshall, T., Slate, J., Kruuk, L., & Pemberton, J. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, *7*, 639–655.

Melo, A. T. d. O., Coelho, A. S. G., Pereira, M. F., Blanco, A. J. V., & Franceschinelli, E. V. (2014). High genetic diversity and strong spatial genetic structure in *Cabralea canjerana* (Vell.) Mart. (Meliaceae): Implications to Brazilian Atlantic Forest tree conservation. *Natureza & Conservação*, *12*, 129–133.

Micheneau, C., Dauby, G., Bourland, N., Doucet, J.-L., & Hardy, O. J. (2011). Development and characterization of microsatellite loci in *Pericopsis elata* (Fabaceae) using a cost-efficient approach. *American Journal of Botany*, *98*, e268–e270.

Morgan, M. T., & Barrett, S. C. (1990). Outcrossing rates and correlated mating within a population of *Eichhornia paniculata* (Pontederiaceae). *Heredity* (Edinburgh), *64*, 271–280.

Ndiade-Bourobou, D., Hardy, O. J., Favreau, B., Moussavou, H., Nzengue, E., Mignot, A., & Bouvet, J. M. (2010). Long-distance seed and pollen dispersal inferred from spatial genetic structure in the very low-density rainforest tree, *Baillonella toxisperma* Pierre, in Central Africa. *Molecular Ecology*, *19*, 4949–4962.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, *89*, 583–590.

Nielsen, R., Tarpy, D. R., & Reeve, H. K. (2003). Estimating effective paternity number in social insects and the effective number of alleles in a population. *Molecular Ecology*, *12*, 3157–3164.

Obayashi, K., Tsumura, Y., Ihara-Ujino, T., Niiyama, K., Tanouchi, H., Suyama, Y., … Muhammad, N. (2002). Genetic diversity and outcrossing rate between undisturbed and selectively logged forests of *Shorea curtisii* (Dipterocarpaceae) using microsatellite DNA analysis. *International Journal of Plant Sciences*, *163*, 151–158.

Oddou-Muratorio, S., & Klein, E. K. (2008). Comparing direct vs. indirect estimates of gene flow within a population of a scattered tree species. *Molecular Ecology*, *17*, 2743–2754.

Palla, F., Louppe, D., & Forni, E. (2002). Sapelli. CIRAD forafri, 4 P.

Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., & Vendramin, G. G. (2005). Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, *14*, 689–701.

Ritland, K. (2002). Extensions of models for the estimation of mating systems using n independent loci. *Heredity* (Edinburgh), *88*, 221–228.

Robledo-Arnuncio, J. J., Austerlitz, F., & Smouse, P. E. (2006). A new method of estimating the pollen dispersal curve independently of effective density. *Genetics*, *173*, 1033–1045.

Robledo-Arnuncio, J. J., Austerlitz, F., & Smouse, P. E. (2007). POLDISP: A software package for indirect estimation of contemporary pollen dispersal. *Molecular Ecology Notes*, *7*, 763–766.

Rousset, F. (2000). Genetic differentiation between individuals. *Journal of Evolutionary Biology*, *13*, 58–62.

Sampson, J. F. (1998). Multiple paternity in *Eucalyptus rameliana* (Myrtaceae). *Heredity* (Edinburgh), *81*, 349–355.

Slatkin, M., & Barton, N. H. (1989). A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, *43*, 1349–1368.

Smouse, P., & Robledo-Arnuncio, J. (2005). Measuring the genetic structure of the pollen pool as the probability of paternal identity. *Heredity* (Edinburgh), *94*, 640–649.

Smouse, P. E., & Sork, V. L. (2004). Measuring pollen flow in forest trees: An exposition of alternative approaches. *Forest Ecology and Management*, *197*, 21–38.

Soularue, J.-P., & Kremer, A. (2014). Evolutionary responses of tree phenology to the combined effects of assortative mating, gene flow and divergent selection. *Heredity* (Edinburgh), *113*, 485–494.

Styles, B. T. (1972). The flower biology of the Meliaceae and its bearing on tree breeding. *Silvae Genetica*, *21*, 175–182.

Tani, N., Tsumura, Y., Fukasawa, K., Kado, T., Taguchi, Y., Lee, S., … Otani, T. (2012). Male fecundity and pollen dispersal in hill dipterocarps: Significance of mass synchronized flowering and implications for conservation. *Journal of Ecology*, *100*, 405–415.

Thomson, F. J., Moles, A. T., Auld, T. D., & Kingsford, R. T. (2011). Seed dispersal distance is more strongly correlated with plant height than with seed mass. *Journal of Ecology*, *99*, 1299–1307.

Vekemans, X., & Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, *13*, 921–935.

Vinson, C. C., Kanashiro, M., Harris, S. A., & Boshier, D. H. (2015). Impacts of selective logging on inbreeding and gene flow in two Amazonian timber species with contrasting ecological and reproductive characteristics. *Molecular Ecology*, *24*, 38–53.

Wang, T., Hagqvist, R., & Tigerstedt, P. M. (1999). Inbreeding depression in three generations of selfed families of silver birch (*Betula pendula*). *Canadian Journal of Forest Research*, *29*, 662–668.

Weis, A. E., & Kossler, T. M. (2004). Genetic variation in flowering time induces phenological assortative mating: Quantitative genetic methods applied to *Brassica rapa*. *American Journal of Botany*, *91*, 825–836.

Weis, A., Nardone, E., & Fox, G. (2014). The strength of assortative mating for flowering date and its basis in individual variation in flowering schedule. *Journal of Evolutionary Biology*, *27*, 2138–2151.

White, G., Boshier, D., & Powell, W. (2002). Increased pollen flow counteracts fragmentation in a tropical dry forest: An example from *Swietenia humilis* Zuccarini. *Proceedings of the National Academy of Sciences*, *99*, 2038–2042.

Wickneswari, R., Ho, W., Lee, K., & Lee, C. (2004). Impact of disturbance on population and genetic structure of tropical forest trees. *Forest Genetics*, *11*, 193–201.

Wright, S. (1921). Systems of mating. III. Assortative mating based on somatic resemblance. *Genetics*, *6*, 144–161.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Monthe FK, Hardy OJ, Doucet J-L, Loo J, Duminil J. Extensive seed and pollen dispersal and assortative mating in the rain forest tree *Entandrophragma cylindricum* (Meliaceae) inferred from indirect and direct analyses. *Mol Ecol*. 2017;26:5279–5291. https://doi.org/10.1111/mec.14241